Polish Infrastructure
for Supporting Computational Science
in the European Research Space

PL-Grid – Status and Plans
The first functioning National Grid Initiative in Europe

www.plgrid.pl/en

Marian Bubak,
Łukasz Dutka, Jacek Kitowski, Zofia Mosurska, Robert Pajak,
Marcin Radecki, Mariusz Sterzel,
Tomasz Szepieniec

ACC Cyfronet AGH, Krakow, Poland

Outline

- Motivation, funding, and objectives
- PL-Grid and EGI
- Organization of the project
- Hardware
- Operational Centre and procedures
- Middleware
- New software and tools
- Training
- Security aspects
- Summary

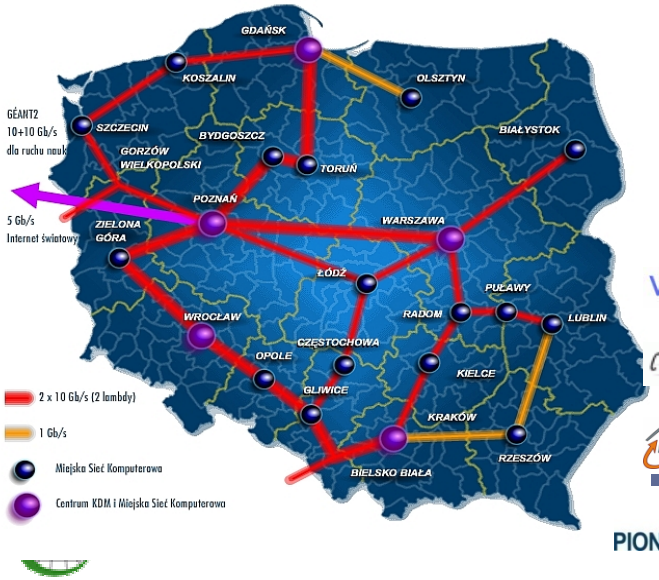
PL-Grid Consortium

- January 2007 - an agreement on creation of the Polish Grid (PL-Grid) Consortium was signed - a response to requirements of Polish scientists and due to ongoing Grid activities in Europe within the framework of EGI_DS.
- Consortium members
  - Academic Computer Centre CYFRONET AGH, Krakow (coordinator)
  - Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University
  - Poznan Supercomputing and Networking Centre
  - Academic Computer Centre, Gdansk
  - Wroclaw Centre for Networking and Supercomputing
- PL-Grid Project proposal which got funded on March 2, 2009.

Rationales behind PL-Grid Consortium

The Consortium consists of five High Performance Computing Polish Centres representing Communities, coordinated by ACC Cyfronet

- Participation in international and national projects
  - ~35 international projects FP5, FP6, FP7 on Grids (50% common)
  - ~15 Polish projects (50% common)
- Polish scientific communities
  - ~75% publications in 5 Communities
- Computational resources
  - Top500 list
- European/Worldwide integration Activities
  - EGEE I-III, EGI_DS, EGI, e-IRG, PRACE, DEISA, OMII, EU Unit F3 „Research Infrastructure” Experts
- National Network Infrastructure available
  - Pionier National Project



Partners' Computing Resources

TOP500 – June 2010

Rank	Site	System	Cores	R_{max} (TFlops)	R_{peak} (TFlops)
161	Cyfronet AGH Krakow	Cluster Platform 3000 BL2x220, L54xx 2.5 Ghz, Infiniband / 2010 Hewlett-Packard	6144	39.93	55.54
181	Gdansk University of Technology, CI Task	ACTION Cluster Xeon E5345 Infiniband / 2008 ACTION	5336	38.17	49.73
444	PCSS Poznan	Cluster Platform 3000 BL 2x220, E5530 2.4 GHz, Infiniband GDR / 2010 Hewlett-Packard	3456	26.22	31.24

PL-Grid Project - Basic Data

- The Project is co-funded by the European Regional Development Fund as part of the Innovative Economy Program.
  - Total budget: 83 M PLN (~ 21 M EUR)
    - Personel cost 27 M PLN (~7 M EUR)
    - Equipment cost 33 M PLN (~8 M EUR)
    - Other cost 23 M PLN (~6 M EUR)
  - Funding from the EC: 68 M PLN (~ 17 M EUR)
- Project duration: 01 January 2009 – 31 December 2011
- Beneficiary: Academic Computer Centre Cyfronet AGH, Krakow, Poland
- Contract number: POIG.02.03.00-00-007/08

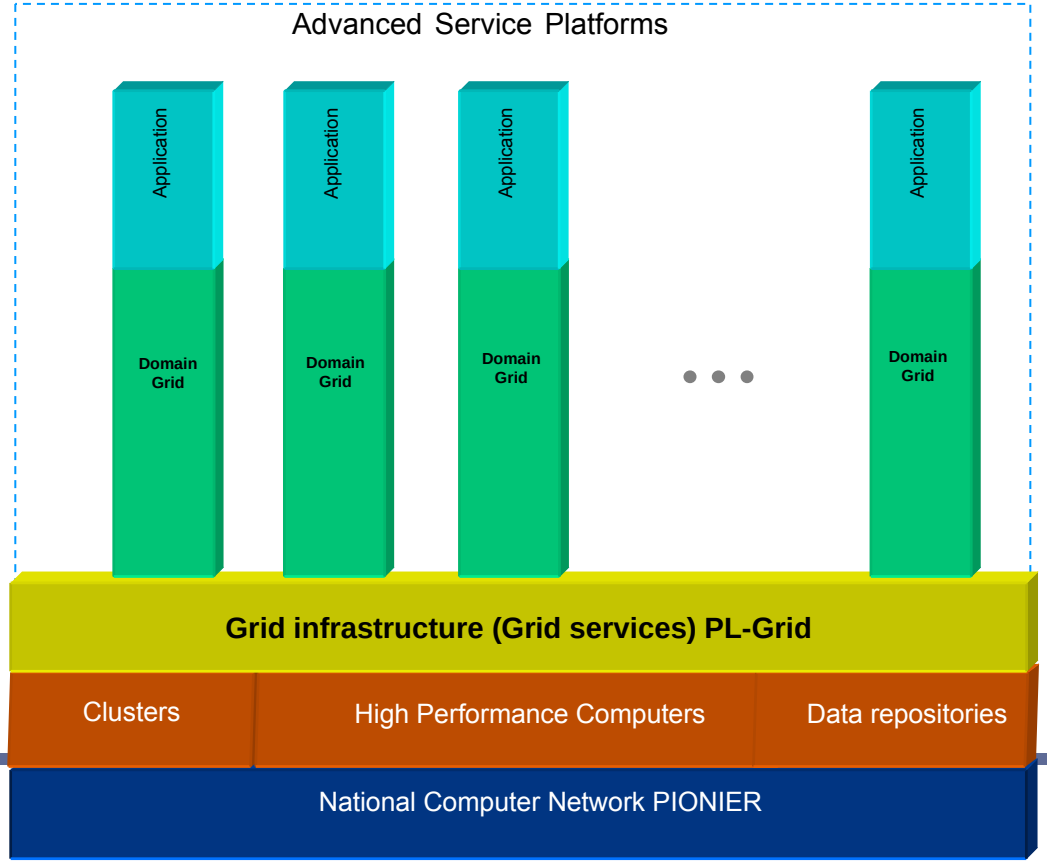
- Project website: www.plgrid.pl/en

Main Objectives of PL-Grid

- Polish Grid is developing a common base infrastructure – compatible and interoperable with European and Worldwide Grids
- Specialized, domain Grid systems – including services and tools focused on specific types of applications
- This approach should enable efficient use of available financial resources
- Plans for HPC and Scalability Computing enabled

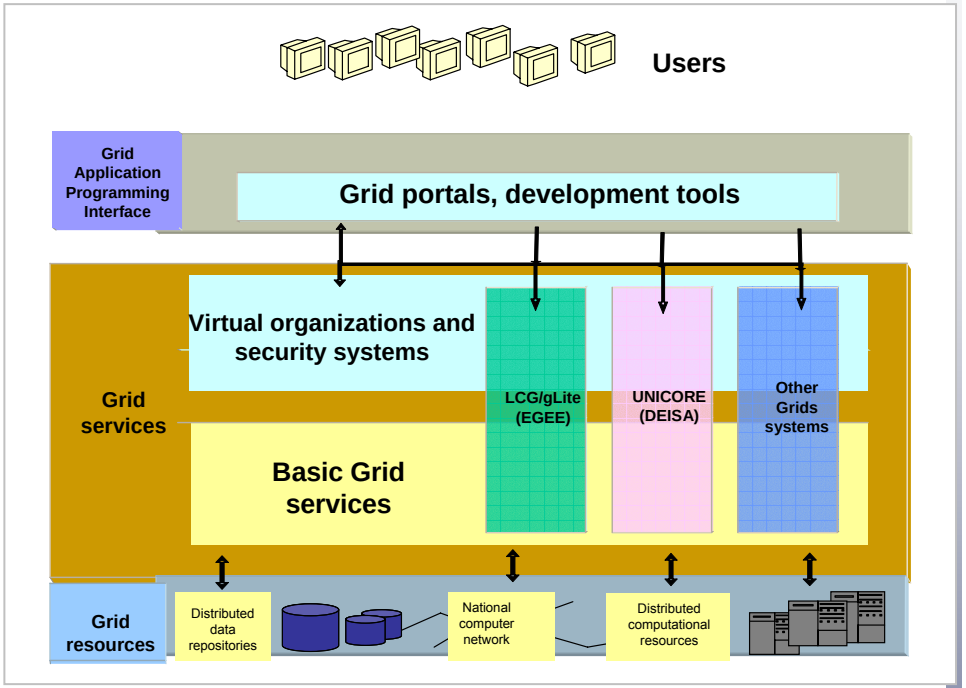
Offer for the Users

- Computing Power 215 Tflop/s
- Storage 2500 TB
- Support from PL-Grid staff on using advanced Grid tools
- Support on porting legacy codes to Grid environment
- Support on designing applications for PL-Grid environment



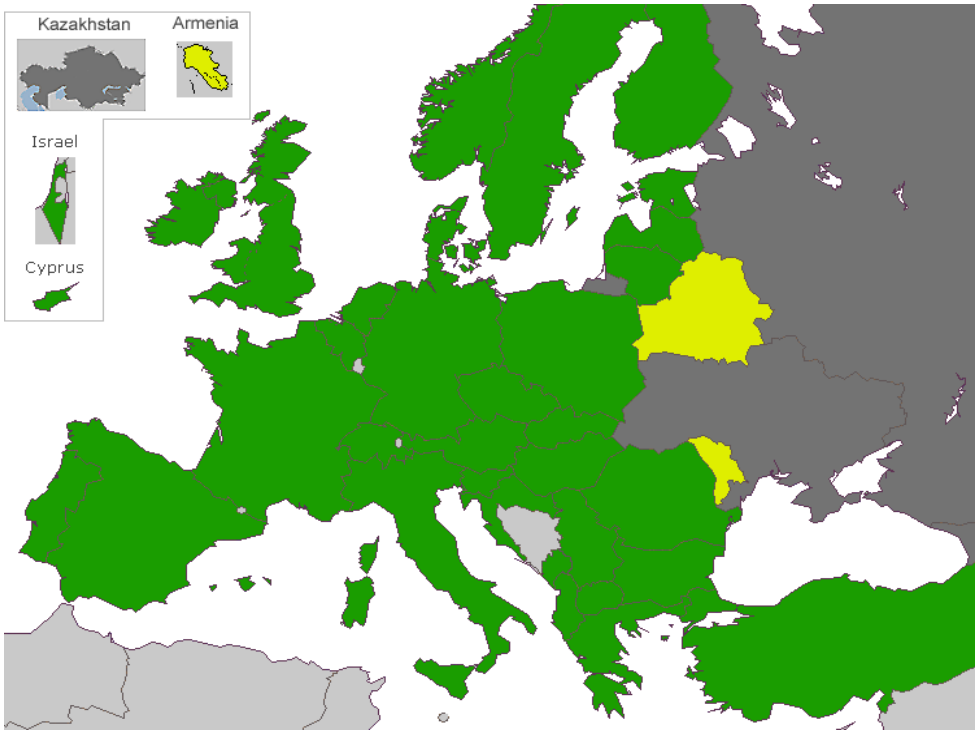
PL-Grid Building Blocks

- PL-Grid software comprises:
  - user tools (portals, systems for applications management and monitoring, result visualization and other purposes, compatible with the lower-layer software used in PL-Grid)
  - software libraries
  - virtual organization systems: certificates, accounting, security, dynamic
  - data management systems: metadata catalogues, replica management, file transfer
  - resource management systems: job management, applications, grid services and infrastructure monitoring, license management, local resource management, monitoring
- Three Grid structures are maintained:
  - production
  - research
  - Development / testing



EGI and PL-Grid

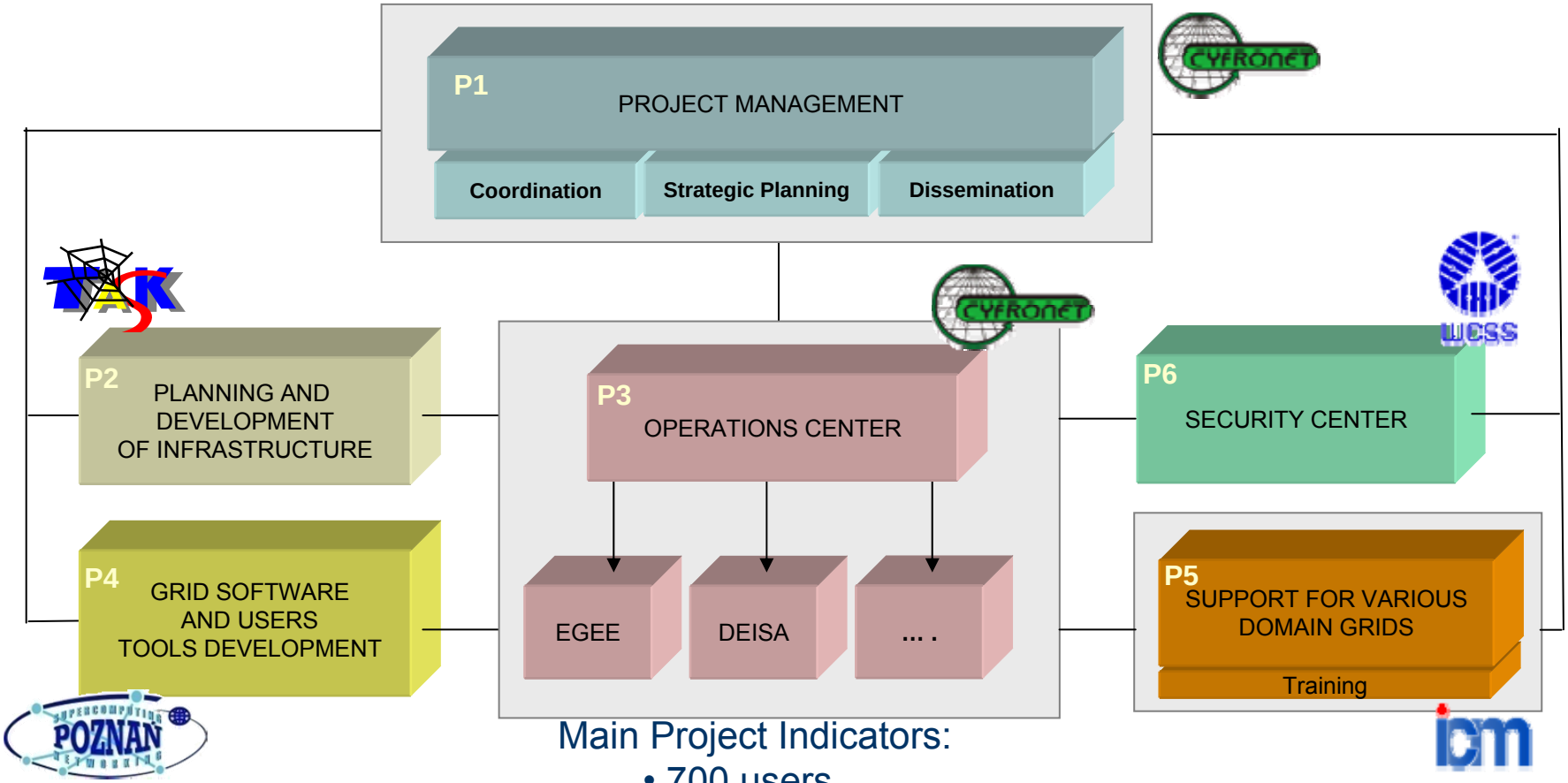
- **EGI.eu** - an organization being developed to coordinate the European Grid Infrastructure, based on the federation of individual National Grid Infrastructures (NGI), to support a multi-disciplinary user community.
- **PL-Grid tasks in EGI**
  - Grid operation and oversight of the e-Infrastructure
  - Coordination of resource allocation and of brokering support for VOs from NGIs



Integration Activity
in the framework of European
Grid Initiative

- to a large extent in the Computational Chemistry – organization and management of Computational Chemistry and Material Science and Technology Specialized Support Centre (CCMST SSC) and EGI liaisons,
- in Development of Unified Middleware via European Middleware Initiative,
- in scientific application porting, especially concerning UNICORE architecture, within Application Porting SSC.

Organization of the PL-Grid project



Main Project Indicators:

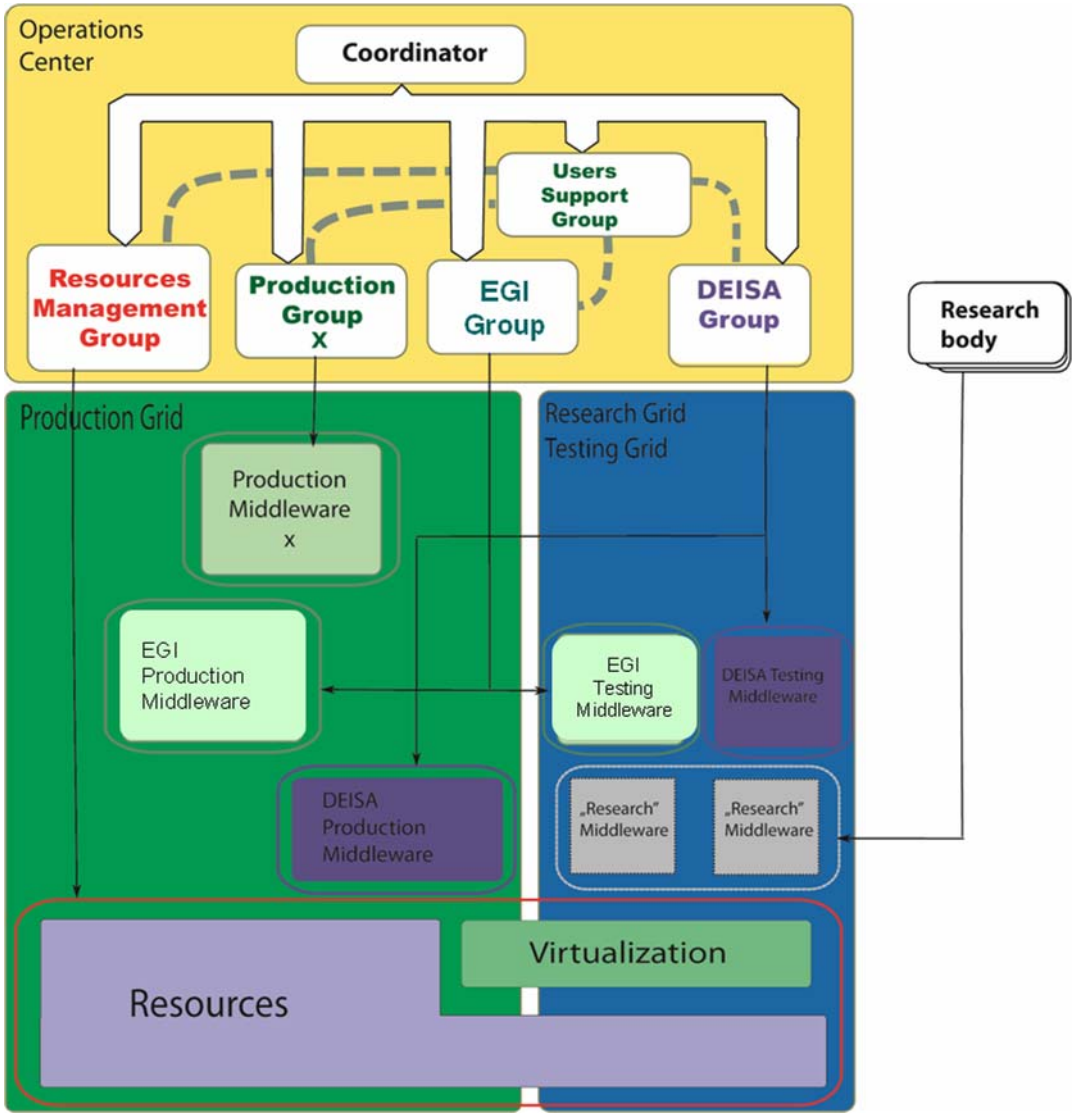
- 700 users
- Peak Perf.: 215 Tflops
- Disk Storage: 2500 TB

Status of Hardware Infrastructure

- Cyfronet, ICM, PCSS
  - power and cooling infrastructure ready for reception of servers and storage
- WCSS
  - final stage of installation of the power and cooling infrastructure
- TASK
  - final stage of tender procedures concerning the power and cooling infrastructure
- Servers and Storage
  - Systems to be deployed in the 1st half of the year:
    - Cyfronet – arbitration finished
    - ICM – arbitration finished
    - PCSS – delivery in progress
  - Systems to be deployed in the 2nd half of the year :
    - WCSS – the tender procedure started
    - TASK – the tender procedure started (in May 2010)
- Plans until the end of 2010:
  - 1900 TB, including: Cyfronet 700 TB, ICM 900 TB, PCSS 300 TB
  - 185 Tflops, including: Cyfronet 43 Tflops, ICM 29 Tflops, PCSS 37 TFlops, TASK 31 Tflops, WCSS 45 TFlops

Tasks of the Operational Center

- Coordination of operations
- Management and accounting
- Collaboration with EGI and PRACE/DEISA
- Users' requirements analysis for operational issues
- Running infrastructure for:
  - Production
  - Developers
  - Research
- Future consideration:
  - Computational Cloud
  - Data Cloud
  - Internal and External Clouds
  - Virtualization aspects



Services of the Operational Center for Users

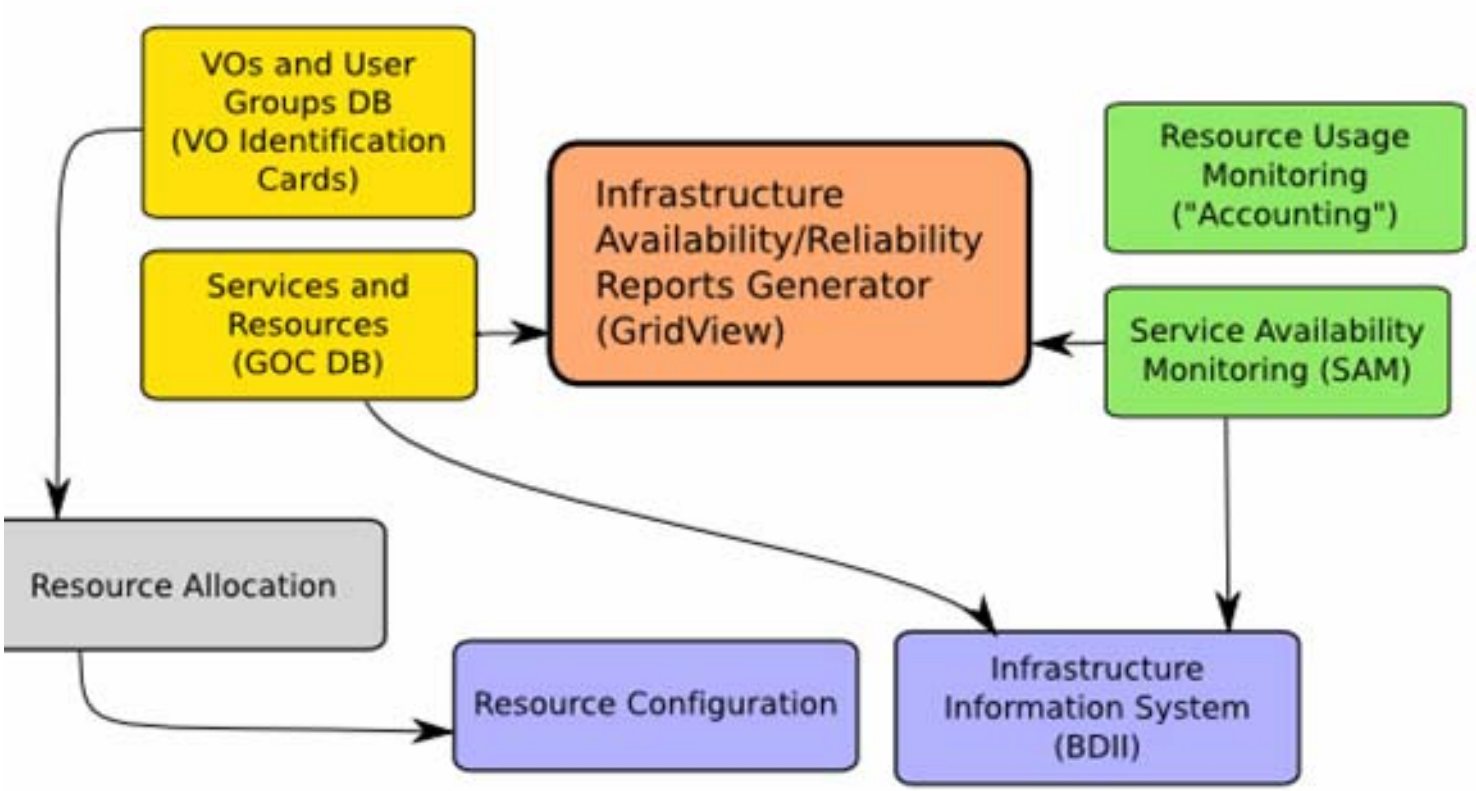
- Operational Center aims at facilitating access to the infrastructure by simplifying the procedures and deployment of useful tools:
  - **System of registration of account management** of the PL-Grid user
    - available at https://konto.plgrid.pl/
    - required entry in the Polish database of „People of Science” or confirmation of the scientific tutor
    - grid access to PL-Grid resources
      - 5 centers – gLite
      - 1 center – UNICORE
    - local access to the queue system
      - „zeus” cluster in ACC CYFRONET AGH
    - ability of application for a grid certificate on-line (soon)
    - application for access to computational services in other centers (soon)
  - **Helpdesk system in PL-Grid**
    - enables reporting and tracking issues
    - available at https://helpdesk.plgrid.pl
    - access also by e-mail: helpdesk@plgrid.pl
    - manual: https://wiki.plgrid.pl/doku.php?id=pakiet5:publiczne:podrecznik_uzytkownika_pl-grid → System Pomocy Helpdesk

Services of the Operational Center for Users - cont'd

- Operational Center cares for a proper functioning of the infrastructure for PL-Grid users by pro-active monitoring of the following infrastructure elements:
  - availability of the infrastructure services
  - software packages supported byPL-Grid
- Provision of the conformity of the PL-Grid and European (EGI) infrastructures
  - software
  - operational procedures
  - security procedures
- Advanced work on the „PL-Grid grants" idea
- Integration of the data presentation concerning resources usage for user
- Work on provision of the integrated user portal

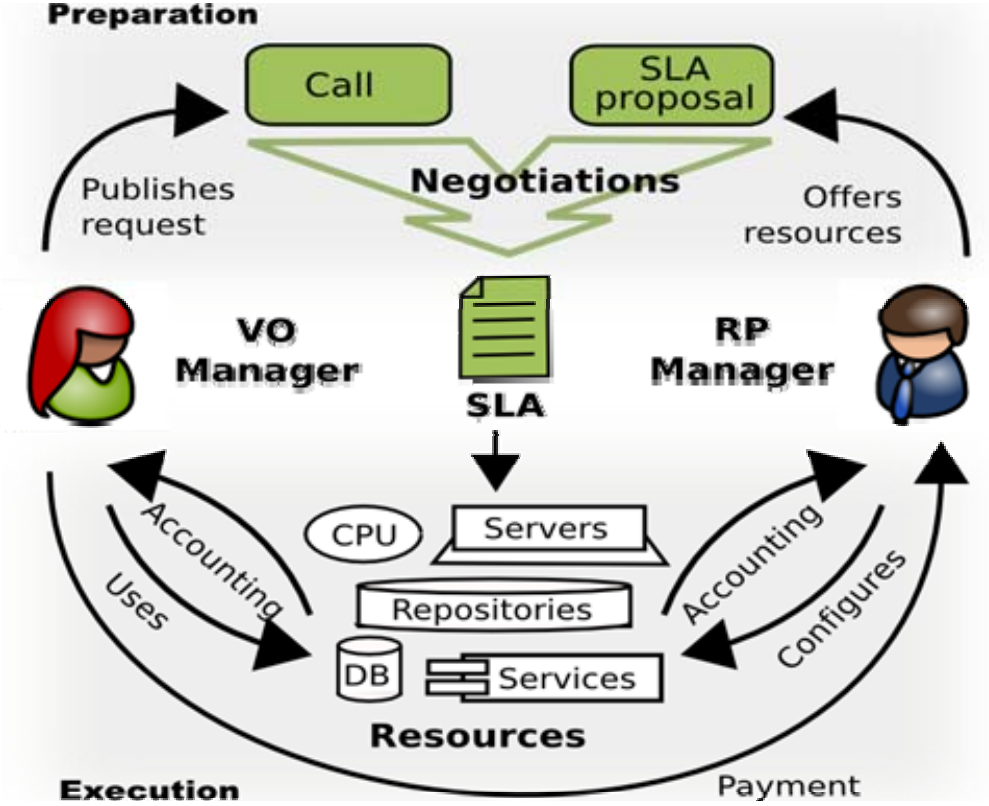
SiteName	NodeName	Status	plg-software-gamess	plg-software-turbomole	plg-software	plg-software-intel
CYFRONET-LCG2	ce.cyf-kr.edu.pl	OK	ok	ok	ok	ok
CYFRONET-LCG2	ce.grid.cyf-kr.edu.pl	OK	ok	ok	ok	ok
PSNC	ce.reef.man.poznan.pl	OK	ok	note	ok	note
TASK	ce.grid.task.gda.pl	OK	ok	note	ok	note
WARSAW-EGEE	ce.polgrid.pl	OK	na	na	crit	na
WCSS64	dwarf.wcss.wroc.pl	OK	ok	ok	ok	note

Operation Model – starting point: EGEE

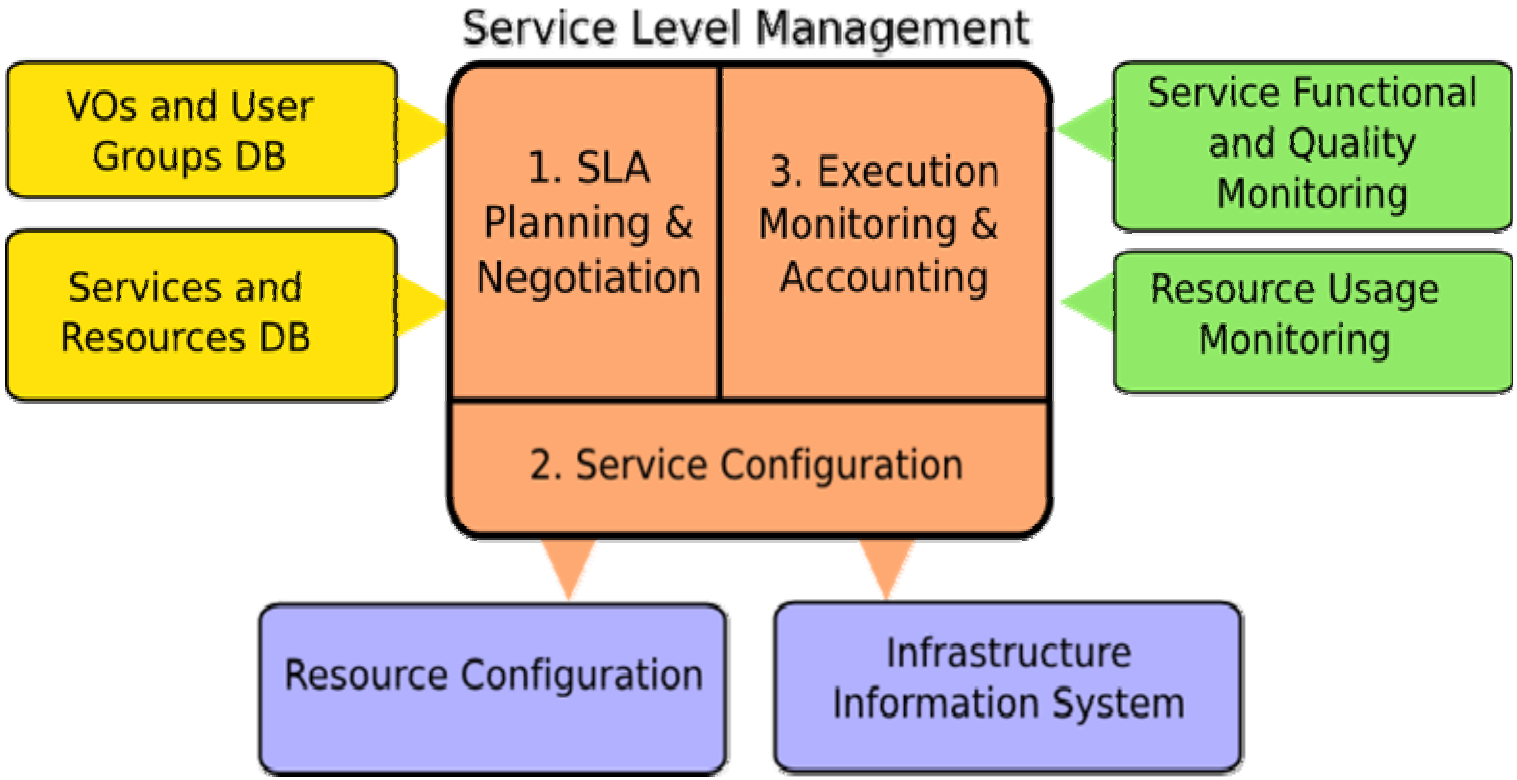


Next Step: Deliver reliable services for users

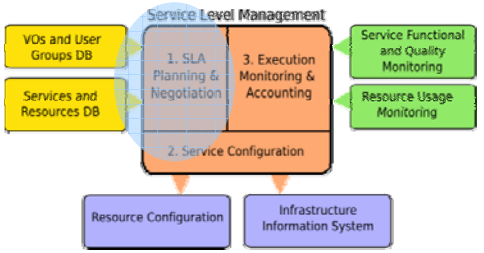
- **Providing resources to users with required qualities of services**
- Required = specified in Service Level Agreement

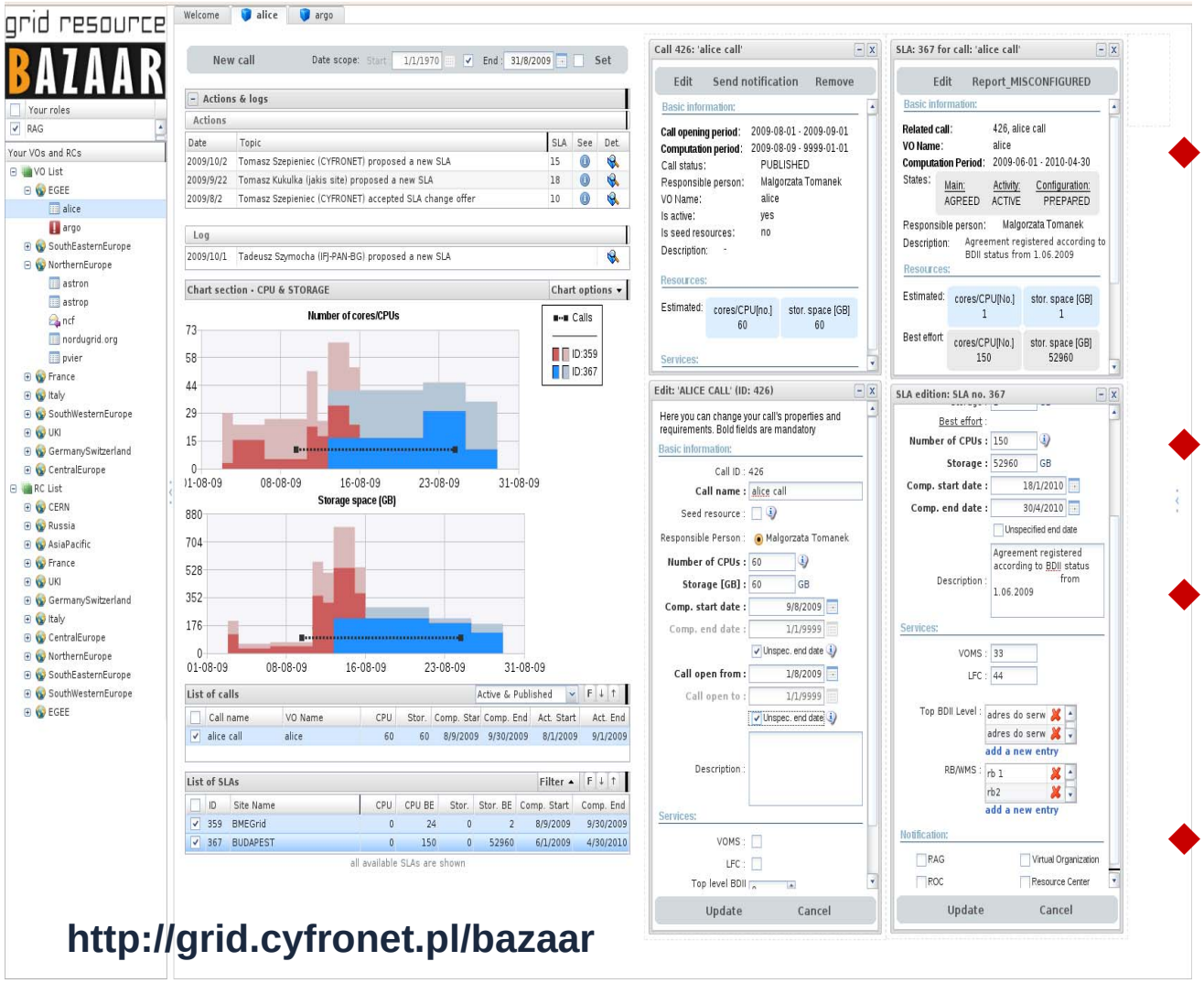


Resource Allocation-related Operation Model



SLA Planning and Negotiation: Tool



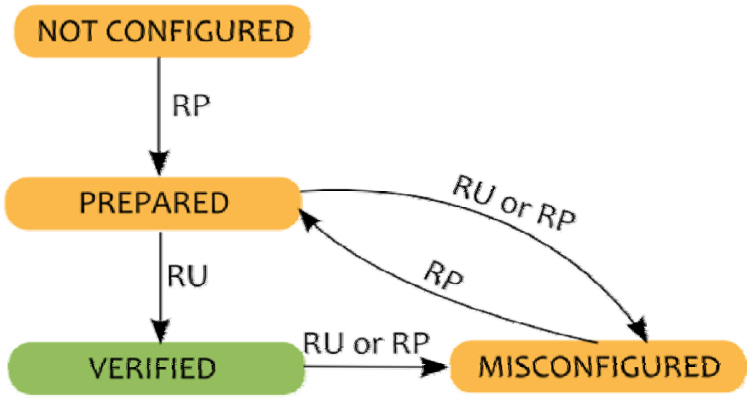


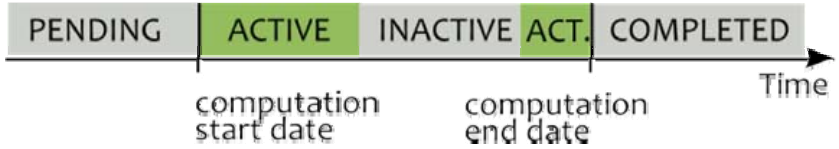
http://grid.cyfronet.pl/bazaar

- Resource Allocation Dashboard for VOs and Resource Providers
- Traceable SLA negotiation process
- V1.2 deployed in CIC Portal used for CE ROC and for seed resources operation
- V2.0 with NGI-role support in alfa testing

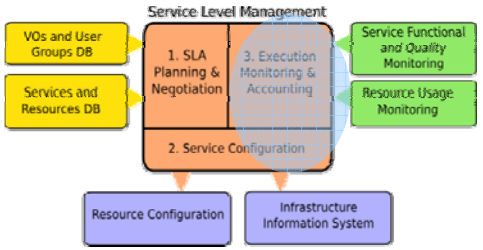
Service Configuration

- Resource/Services are configured according to the SLA:
  - Limits
  - Priorities
  - Reservations
  - Quotas
  - Software required
  - ....
- Verification of a site configuration by a VO is required
- Only sites having an agreed, active SLA with a VO with verified configuration are available in Infrastructure Information System
  - this prevent not-verified resources to be put into 'production'

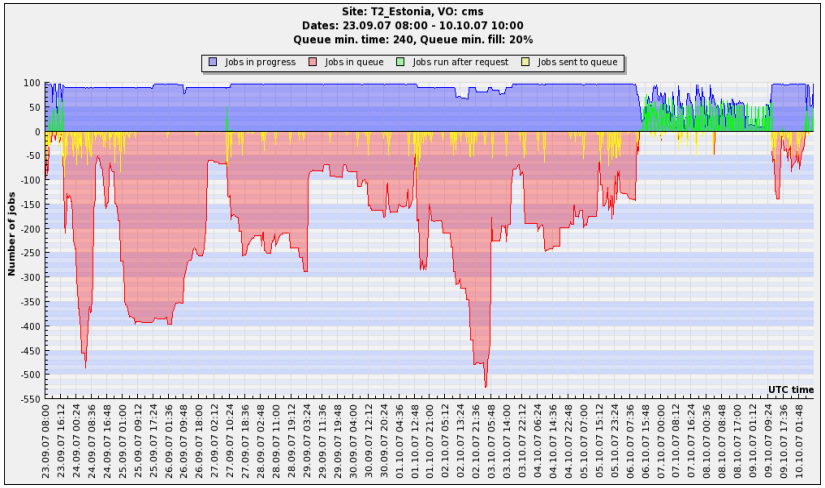




SLA Monitoring



- Infrastructure monitoring results are used to monitor SLA metrics
  - Critical tests
  - Availability/reliability
  - ....
- Accounting data are used to verify SLA metrics:
  - Resource level
  - Failed job ratio
  - Waiting time
  - ....
- Option to suspend a site that breaks its SLA
- Feedback about sites/VOs can be published

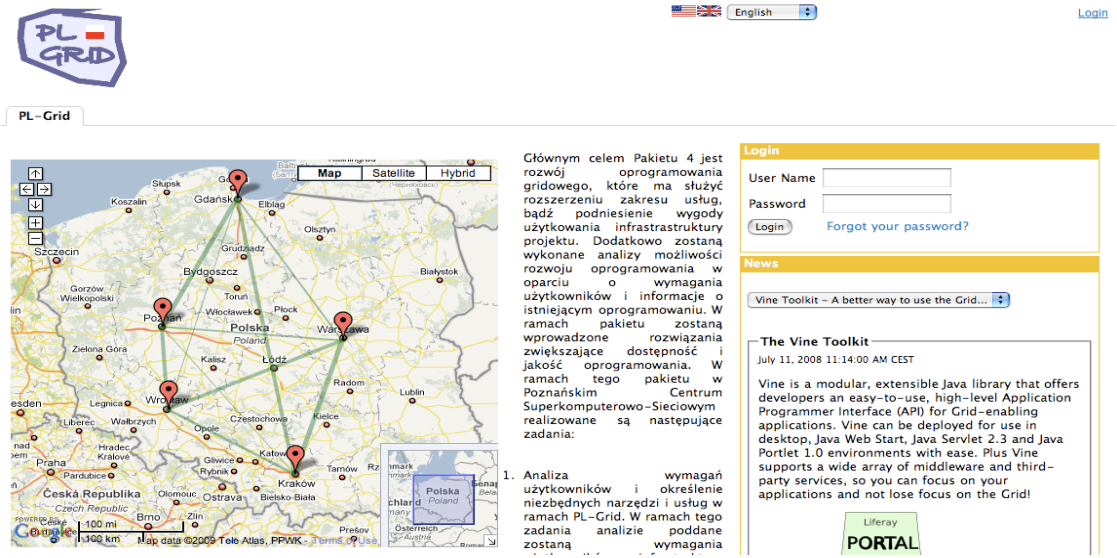


Software packages

- Access to software packages will be provided by:
  - gLite
  - UNICORE
- Examples of available packages:
  - In the field of biology: AutoDock, BLAST, ClustalW2, CPMD, Gromacs, NAMD
  - In the field of quantum chemistry: ADF, CFOUR, Dalton, GAMESS, Gaussian, Molcas, Molpro, MOPAC, NWChem, OpenBabel, Siesta, TURBOMOLE
  - In the field of physics: ANSYS FLUENT, Meep
  - In the field of numerical computations and simulation: Mathematica, MATLAB
  - Other: Blender, POV-Ray
- If needed, most of the packages - available until now only on the High Performance Computers in several Polish computing centres - can be made available, as well as new software suggested by users
- Users may report us their expectations through a survey available at: http://www.plgrid.pl/ankieta
- The system of testing the software packages in the grid environment has been prepared and deployed
- The correctness of functioning of the packages is monitored automatically in each of the centers

New software and services (1/3)

- Close cooperation of 8 programming and testing groups, about 20 people
- Installation and provision for testing purposes gLite, Unicore and QosCosGrid
- About 30 various configurations of virtual machines with installed software used for development and testing of the tools for users – the choice of the technology made
- Functional, conformity and efficiency tests of selected packages of the research software made in order to perform the deployment and support of the new tools and services on the production level

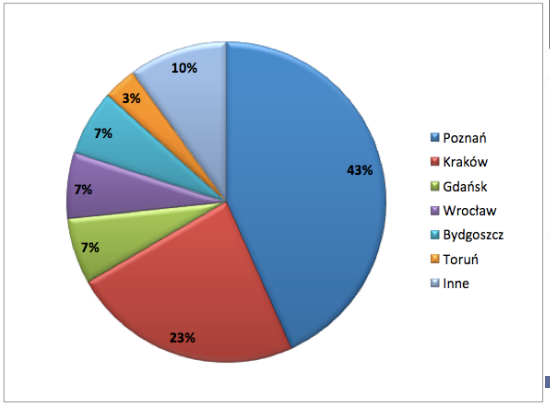


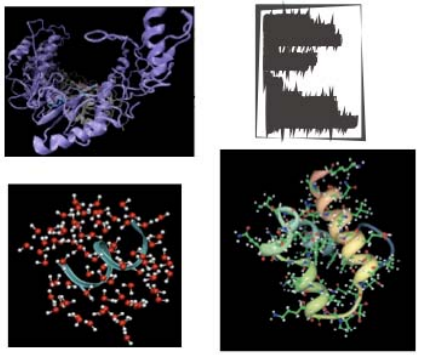
New software and services (2/3)

- Extension of the GridSpace platform with the set of new functions, support for new scripts and integration with new grid services
- Implementation and provision of the advanced graphical interfaces, visualization and tasks and data management for selected applications of the new users by use of the Vine Toolkit
- Efficiency and functional tests of the middleware services QosCosGrid and integration with gLite and Unicore infrastructure at the queue systems level
- Integration of the Migrating Desktop and gEclipse tools with various middleware services in PL-Grid
- Plan of extension and deployment of the new tools FiVO for management and monitoring of the virtual organizations
- Test versions of the tools for users and systems administrators: Bazaar, mobile access to the infrastructure, new security applications
- Integration of the selected tools and web applications with Liferay portal framework and Nagios monitoring system

New software and services (3/3)

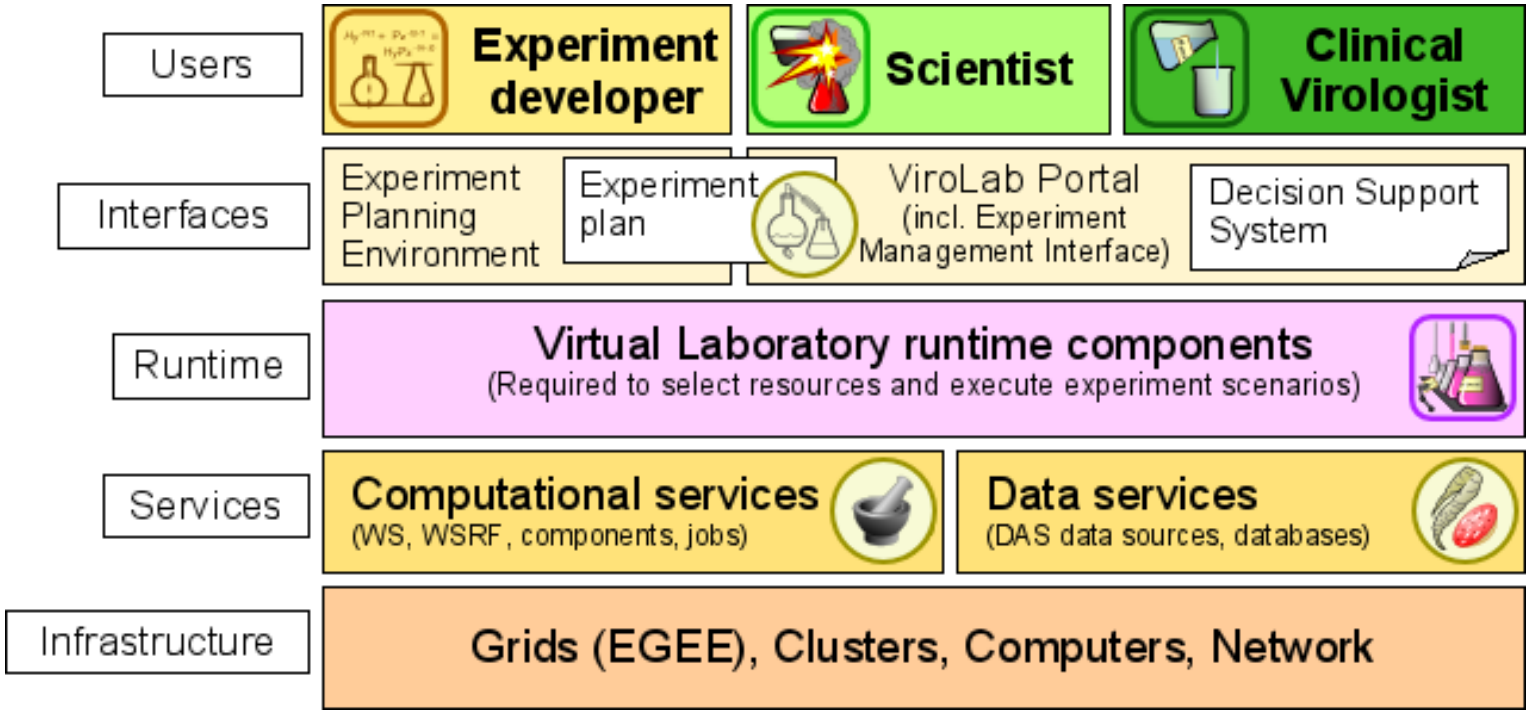
- Direct contact with new users on the basis of a survey, available at: **www.plgrid.pl/ankieta**
- Requirements of the Polish users (results ~100 surveys) considered in the new applications, tools and services developed and tested in the framework of the Package 4
- Large group of users cooperating with the software and tools team
  - Department of Chemistry of the Jagiellonian University
  - Department of Bioinformatics and Telemedicine of the Collegium Medicum of the Jagiellonian University
  - University of Adam Mickiewicz
  - Poznan Technical University
  - Wrocław Technical University
  - Administrators of the computing centers
  - …





Example: Virtual Laboratory GridSpace

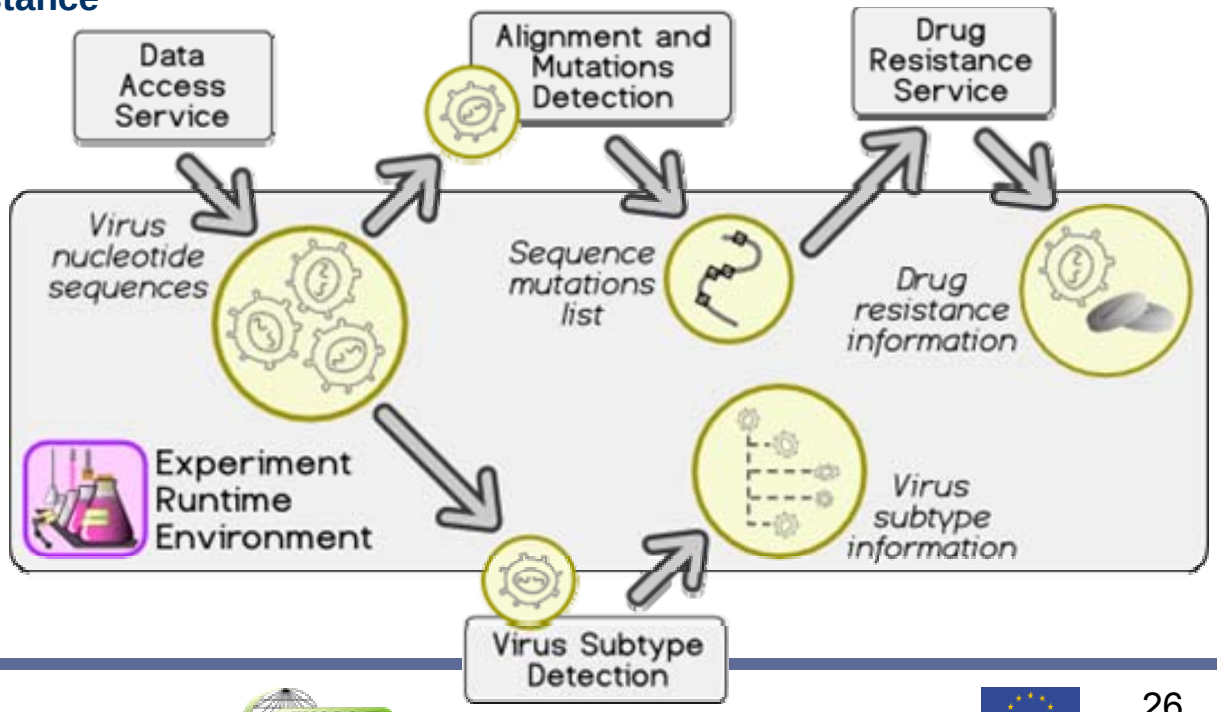
- Use of distributed computational resources and data repositories
- High-level tools offered for the user for in-silico experiments



Sample experiment in ViroLab Environment

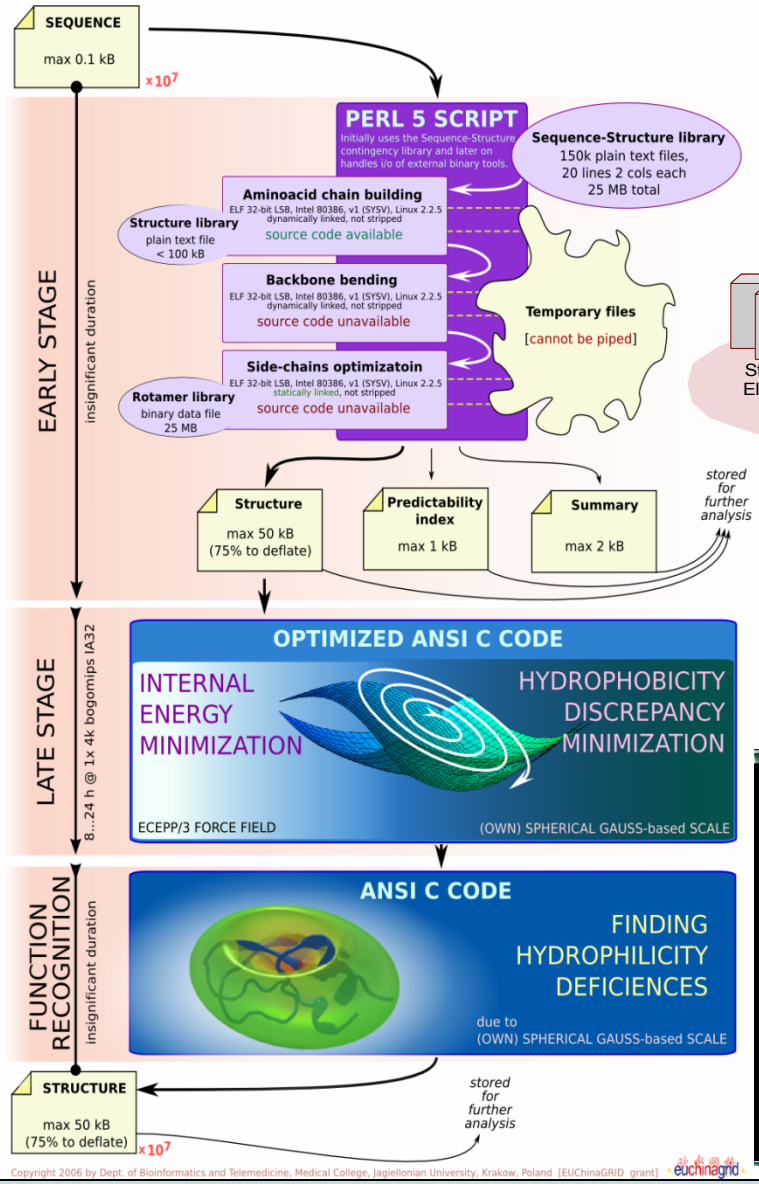
- Patient's data
  - **Medical examination**
  - HIV genetic sequence put into database
- Experiment *in-silico*
  - **Collect HIV genetic sequences from database**
  - **Perform sequence matching**
  - **Calculate virus resistance**

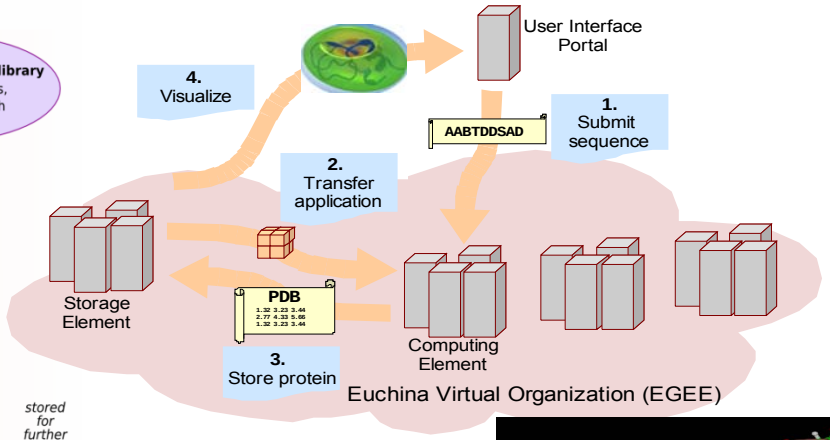
http://gs2.cyfronet.pl/
http://www.virolab.org

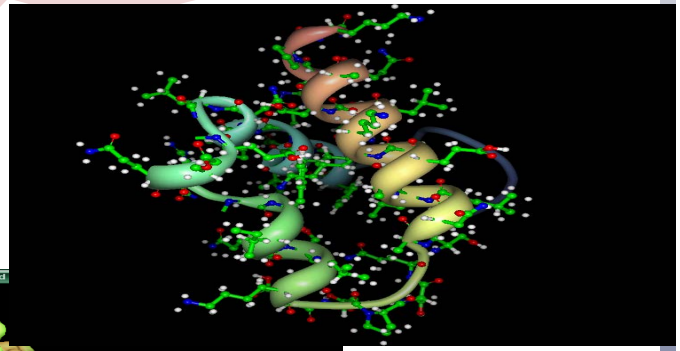


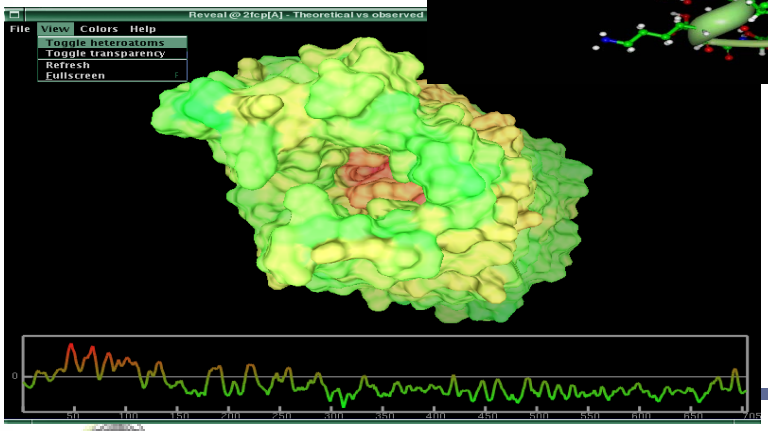
Example: Biotechnology in Grid

Never Born Protein Folding



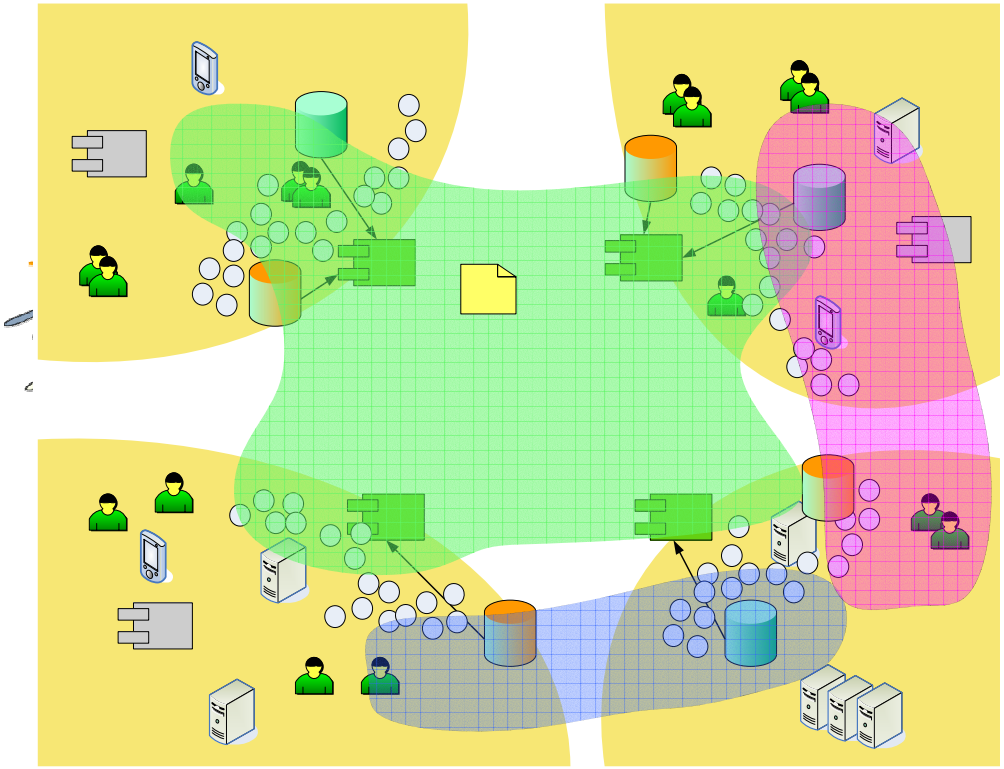






ACK: Irena Roterman, Jagiellonian University, Tomasz Szepieniec, Cyfronet

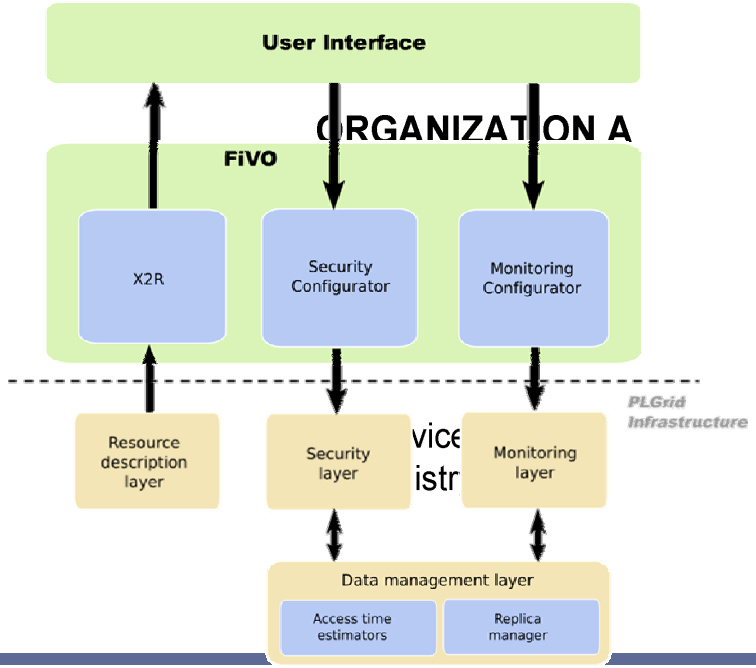
Contract-based Dynamic Virtual Organizations FiVO



Goal:

- Allow end users to defined their requirements for the Virtual Organization on a high level of abstraction
  - Semantic description of domain

Overall system architecture



Results:

- Automatic deployment of the VO
  - Security and monitoring
- Optimization of data access
  - Replication and migration

MDS

Training and Users' Support

- Basic training on access to the infrastructure through gLite and UNICORE conducted in all centers participating in the project – in Gdańsk / Kraków / Poznań / Warszawa / Wrocław
- More advanced training started
- Similar (free) training may be conducted in other centers, if necessary
- eLearning training will be available soon (prepared by use of Blackboard system)
- Helpdesk system implemented
  - it's a novel support system for people using the Project resources
  - it involves the technical support and organization of the current users' support by the experts (maintenance of trouble tickets)
  - tickets in Helpdesk may be created by sending an email to: helpdesk@plgrid.pl
  - Online system, available at: https://helpdesk.plgrid.pl

Security in PL-Grid

- Provision of two CAs – PKI certification centers – for grid users
- Project and implementation of the SimpleCA system, facilitating the users obtaining PKI certificates and their usage
- Project and implementation of the secure configuration of the infrastructure, in conformity with the most actual security standards
- Project of the system monitoring the conformance of the configuration deployed in the centers with the security policy
- Creation of the group of experts from the field of security, in order to continuously monitor the environment, immediate react on incidents and support users and administrators
- Prototype version of the system of correlation of information about the attacks on the infrastructure (ACARM-ng)
- Audits of applications crucial for grid security

Summary of Activities

- Achieved
  - Development of provided services
  - Provision of resources for covering operational costs and international cooperation
  - Start of cooperation with EGI.eu, D-Grid and BE-Grid
- Long term
  - Software and tools implementation
  - Users' support and traning
  - Provision, maintenance and extension of the necessary infrastructure
  - Development and implementation of new computational paradigms and environments integration
    - HPC and distributed computing (HPCaaS, IaaS, PaaS, SaaS)
    - National Cloud Initiative (computing clouds, data clouds)
    - SOA paradigm, knowledge usage …
    - „Future Internet" as defined by EC in Workprogramme
- Strategical
  - Development of the domain specific environments

Acknowledgements

- ACC Cyfronet AGH
  - Jacek Kitowski
  - Tomasz Szepieniec
  - Marcin Radecki
  - Mariusz Sterzel
  - Agnieszka Szymańska
  - Zofia Mosurska
  - Andrzej Oziębło
  - Tadeusz Szymocha
  - Aleksandra Mazur
- ICM
  - Piotr Bała
  - Maciej Filocha
- PCSS
  - Norbert Meyer
  - Krzysztof Kurowski
  - Mirosław Kupczyk
- WCSS
  - Józef Janyszek
  - Bartłomiej Balcerek
  - Paweł Dziekoński
- TASK
  - Mścisław Nakonieczny
  - Jarosław Rybicki
  - Rafał Tylman

http://www.plgrid.pl/en

CGW'10


This year CGW celebrates its 10th anniversary!
Cracow '10
Grid Workshop
Kraków, Poland
October 11-13, 2010
Topics
e-Science, system-level science and collaborative applications,
models, methods and tools for collaborative applications development,
virtual laboratories and problem solving environments,
distributed computing infrastructures, grids and clouds,
knowledge in e-Science and DCI systems,
virtual organizations and security aspects,
resource management and scheduling,
monitoring and information management,
software engineering aspects,
industrial and social implications.
Keynote Speakers
Kyriakos Baxevanidis, European Commission, Brussels, Belgium
Ewa Deelman, University of Southern California, Los Angeles, USA
Ad Emmen, AlmereGrid and EnterTheGrid, The Netherlands
Cees de Laat, Universiteit van Amsterdam, The Netherlands
Jarek Nabrzyski, University of Notre Dame, USA
Steven Newhouse, EGI.eu and EGI-InSPIRE Project
Dana Petcu, Western University of Timisoara, Romania
Deadlines
abstract submission – Sept. 05, 2010
acceptance notification – Sept. 19, 2010
early registration – Sept. 26, 2010
www.cyfronet.pl/cgw10/
Sponsors
Organized by:
CYFRONET
Academic Computer Centre Cyfronet AGH
AGH
Institute of Computer Science
Institute of Nuclear Physics PAN
INNOVATIVE ECONOMY

FGCS

- Perspectives on grid computing by Uwe Schwiegelshohn + 20